

AI-Driven Stock Market Forecasting Using Financial Statement Analysis

By

Marwa Othman Esaa

Midocean University

Advisor

Dr. Ashraf Elsayed

Abstract

Using machine learning for stock prediction has gained a lot of interest recently. Many studies show that machine learning algorithms can forecast stock movements effectively by using historical data. However, most current methods mainly focus on short-term predictions based on past stock prices and technical indicators, such as the Relative Strength Index (RSI) and Moving Average Convergence Divergence (MACD). In this study, we take a different approach. We use 41 quarters of financial statement data to explore the possibility of predicting long-term stock returns based on fundamental indicators. The model includes important financial ratios, such as profitability, leverage, efficiency, liquidity, and market value ratios, to create a predictive framework. We trained three machine learning algorithms: Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Random Forest (RF).

Introduction

Background and Motivation

The main reason for predicting stock price movements is the chance for profit. Since financial markets began, investors have looked for ways to forecast stock performance to find profitable opportunities. This involves buying stocks expected to increase in value and selling them for a profit. Over the years, many stock prediction methods have appeared, but their consistency and reliability are still debated.

Stock prediction methods can generally be grouped into three main categories:

Fundamental Analysis. This approach predicts stock prices by looking at the financial health of companies using publicly available data, including financial statements and economic indicators. It aims to estimate a stock's true value.

Technical Analysis. This technique relies only on past market data such as prices and trading volumes. In contrast to fundamental analysis, technical analysis assumes that all public information is already included in a stock's price. It is based on three main ideas:

- Market action discounts everything.
- Prices are moving in trends.
- History repeats itself.

Sentiment Analysis. This method assesses investor sentiment by examining news articles, analyst reports, and online commentary related to specific stocks or the overall market. It seeks to identify the current mood or emotional tone influencing investor behavior.

Among these methods, technical analysis and sentiment analysis are mainly used for short-term predictions, for days or even hours, while fundamental analysis is usually applied for mid- to long-term forecasts, spanning quarters or years. Recently, combining machine learning (ML) and data mining techniques into financial modeling has gained significant attention. Many existing ML-

based stock prediction studies focus on short-term forecasting using technical indicators or sentiment data. However, these models often neglect real-world factors like transaction costs, which can affect the practical use and reliability of their results.

This research aims to fill this gap by examining machine learning methods for long-term stock prediction based on fundamental analysis. The study specifically investigates the predictive power of financial statement data in estimating future stock performance. To achieve this, we compare the prediction performance of three machine learning models trained on fundamental financial features. Our dataset includes quarterly financial data from 100 companies listed on the TASI (Tadawul All Share Index) between 2015 and 2025. We assess model performance by ranking 70 stocks based on their predicted relative returns and building portfolios accordingly. The actual returns of these portfolios serve as the main measure for evaluating prediction accuracy and model effectiveness.

Problem Statement

Despite the increasing use of artificial intelligence in financial markets, most stock prediction models focus mainly on technical analysis. They often overlook the valuable insights found in corporate financial statements. Traditional machine learning models also have difficulty capturing complex, non-linear relationships between fundamental financial indicators and stock price behavior. This research aims to fill this gap by creating a machine learning-based model that predicts stock trends using fundamental financial data. Specifically, the study looks at how key profitability ratios, such as Return on Assets (ROA), Return on Capital (ROC), Return on Equity (ROE), and Return on Common Equity, affect stock price movements. The main goal is to improve the accuracy, clarity, and reliability of long-term stock price predictions by using AI-driven methods.

Research Objective

This study proposes a new method for estimating the true value of stock return. It uses information from financial statements and applies machine learning techniques. The goal is to improve the decision-making abilities of investors and financial analysts. This approach offers a data-driven way to evaluate stocks for the long term.

METHODOLOGY

Data Preparation

The sample in this study was collected from the firms included in Tadawul All Share Index (TASI) which has more than 200 companies listed on its main market. To ensure the study's relevance, attention was given to companies within the strict core industries of the index. For these companies, historical overtime data was available on a quarterly basis from the first quarter of 2015 to the first quarter of 2025, and this data was collected using web scraping techniques from publicly available data.

The original data set had a wide range of both financial and non-financial variables, but it was also a dataset with a lot of missing values. A large part of the missing values formed large clusters for some features and other missing values were randomly placed in the dataset. A combination of feature removal and data imputation techniques were found to be adequate to resolve these missing values.

We removed features that had over 50 percent missing data. Moreover, non-fundamental features such as stock high and low prices were removed to sharpen the focus on actual quantitative fundamentals. Even after the above steps were taken, the remaining features still contained some sparse missing values. These were resolved using linear interpolation with adjacent quarter values. If, for example, the ROE for Q3 2015 was absent, we would estimate it by taking the 2Q and 4Q values for 2015, averaging them, and assigning that as the ROE for Q3 2015.

After completing the data preparation procedures, the resulting dataset included 70 companies, each with 41 quarterly records over 10 years, yielding a total of 2,870 records. For further analysis, 13 financial features were extracted, which are detailed in Table 1.

Return on Assets	
Return on Capital	
Return on Equity	
Return on Common Equity	
Gross Profit Margin	
EBITA Margin	
EBIT Margin	
Net Income Margin	
Asset Turnover	
Current Ratio	
Debt / Equity	
Long Term Debt / Capital	
Book Value Per Share	

Training

In this lesson, we tried out two different training techniques: constructing a single model to predict returns for all stocks and constructing a model for each stock individually. These two methods are often referred to as global learning and local learning. Global learning is when a single model is created to predict all stock market data, offering a wider scope of data to understand relations between various stocks. In comparison, local learning builds individual models for each stock, allowing for bespoke learning tailored to that stock's profile. Global learning might have the advantage of a larger training dataset, while local learning usually outperforms global learning in capturing specific stock patterns richer in detail. In initial tests, the local learning method outperformed every other in the local setting. For this reason, we embraced local learning as a strategy, creating individual models for every stock using all three selected machine learning algorithms.

Evaluation Metrics

The goal of this project is to develop a system which can be used to guide stock portfolio design strategy for long term

investment. Therefore, simple and general evaluation methods are preferred. We decided to build regression models to predict the price for each stock and then induce a ranking of the stocks by sorting their predicted relative returns. The ranking can then be used for portfolio design, and the actual performance of the portfolios in terms of real relative return can be evaluated with ease.

MAE

Top-K

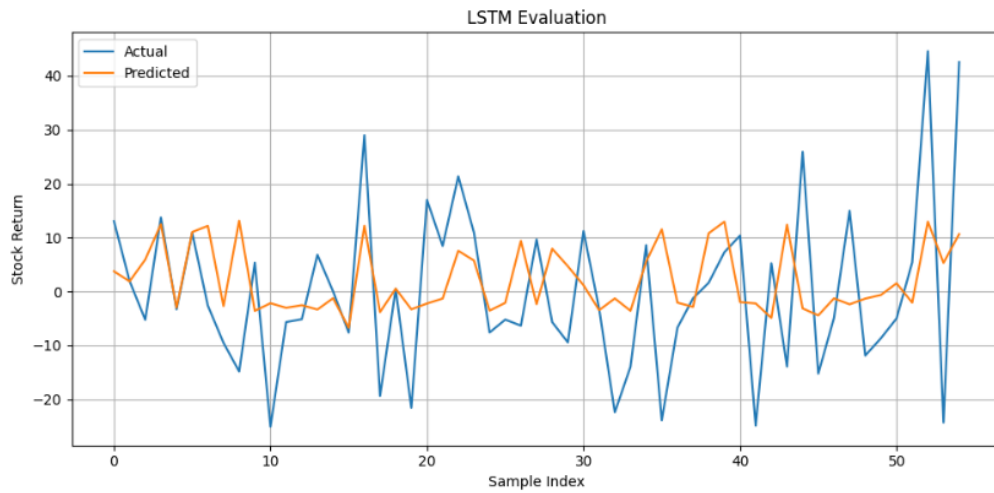
Portfolio Score

RESULTS

LSTM HYPERPARAMETERS

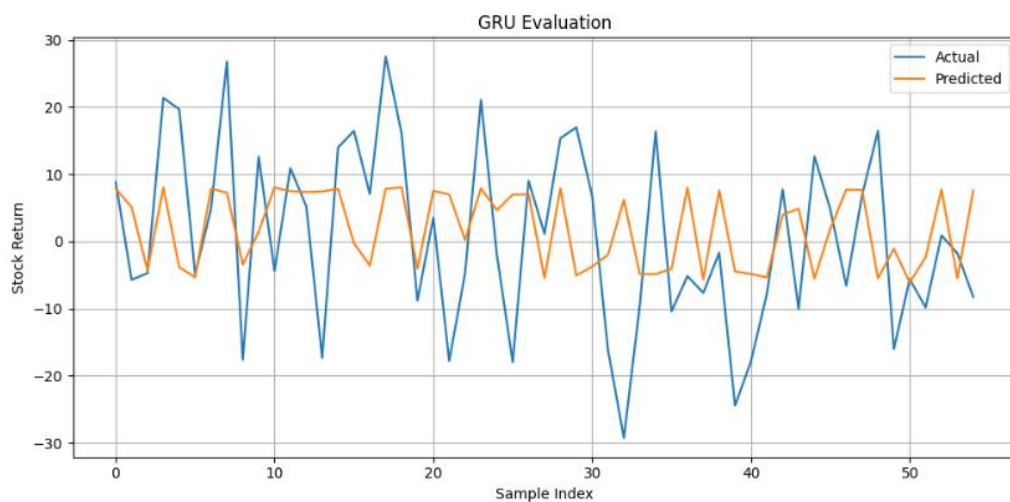
Wind_size	Number of inputs	hidden size	Number of LSTM layers	Dropout rate	Learning rate	Batch size	epochs
20	13	64	2	0.4	0.0005	128	70

Validation MAE: 11.6773



GRU HYPERPARAMETERS

Wind_size	Number of inputs	hidden size	Number of LSTM layers	Dropout rate	Learning rate	Batch size	epochs
16	13	64	2	0.4	0.0005	128	70



Evaluation Metrics MAE

Model	MAE
LSTM	11.6773
GRU	10.235